(4)

AD-A220 364

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER #54	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) An Averaging Result for Random Differential Equations		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Nathan Intrator		8. CONTRACT OR GRANT NUMBER(s) N00014-86-K-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Neural Science Brown University Providence, Rhode Island 02912		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS N-201-484
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel & Training Rsch. Program Office of Naval Research, Code 442PT Arlington, Virginia 32217		12. REPORT DATE April 12, 1990
		13. NUMBER OF PAGES 13 pages
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited. Publication in part or in whole is permitted for any purpose of the United States Government.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DTIC
S ELECTE D
APR 12 1990
B

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Random differential equations
Stochastic approximations
Random evolution
Neural networks

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper concerns differential equations which contain strong mixing random processes. The solution process is shown to be well approximated by a deterministic trajectory, over an infinite time interval, using the interplay between the rate of fluctuations of the random process and the rate of the ϕ mixing. An application of the result is given for analyzing synaptic modifications in Neural Networks.

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

90 04 12 134

AN AVERAGING RESULT FOR RANDOM DIFFERENTIAL EQUATIONS

Nathan Intrator

Division of Applied Mathematics and
Center for Neural Science
Brown University
Providence, Rhode Island 02912
Email: nin@brownvm.bitnet

March 9, 1990

Abstract This paper concerns differential equations which contain strong mixing random processes. The solution process is shown to be well approximated by a deterministic trajectory, over an infinite time interval, using the interplay between the rate of fluctuations of the random process and the rate of the φ mixing. An application of the result is given for analysing synaptic modifications in Neural Networks.

1. Introduction

The mathematical theory of stochastic differential equations is concerned mainly with the study of Itô equations and the associated Markov process. Mostly, the results on non Itô type equations have been concerned with the conditions under which $x_\epsilon(t)$ converges (as $\epsilon \to 0$) to a diffusion process on finite intervals $[0, T/\epsilon]$ (cf. Stratonovich, 1963; Cogburn and Hersh, 1973; Papanicolaou and Kohler, 1974; Blankenship and Papanicolaou 1977). Averaging results for random differential equations are usually discussed in conjunction with the law of large numbers Kohler and Papanicolaou (1976) with the central limit theorem for $(x_\epsilon(t) - y_\epsilon(t))/\sqrt{\epsilon}$ on $[0, T]$ (cf. Khasminskii, 1966; and White 1976). Geman (1979) showed that the solution process of a random differential equation which contains strong mixing random process is well approximated by a deterministic trajectory over a finite time interval, and for a more restricted systems, over the infinite time interval. Analysis analogous to that was carried out on Itô type equations by Vrkoc (1966), and by Lybrand (1975).

In this paper we shall continue the direction taken by Geman and approximate the solution process by a deterministic trajectory over an infinite time interval, using the interplay between the rate of fluctuations of the random process and the rate of the φ mixing, yielding a result for a wide family of nonlinear random differential equations. We will establish conditions under which the random solution *stays close* in L^2 sense to the associated deterministic solution. The result is particularly useful when a converging deterministic equation is approximated by a random equation that is more computationally feasible. Section 4 is devoted to such an application, in the theory of synaptic modification in Neural Networks.

Similar analysis was carried out on the discrete time version of such equations, see Ljung (1978), Kushner and Clark, (1978), Dupuis and Kushner (1987), and the references therein.

2. Formulation and statement of the problem

In this section we briefly summarize the relevant results form Geman (1977, 1979).

Let $\phi(t,\omega)$ be a bounded stationary stochastic process with $\mathcal{F}_0^t$ and $\mathcal{F}_t^\infty$ the σ-fields generated by $\{\phi(\tau,\omega) : 0 \le \tau \le t\}$, and $\{\phi(\tau,\omega) : t \le \tau < \infty\}$ respectively. Let the signed measure $v_{t,\delta}$ be defined on $(\Omega \times \Omega, \mathcal{F}_0^t \times \mathcal{F}_{t+\delta}^\infty)$ by

$$v_{t,\delta} = P(\omega : (\omega,\omega) \in B) - P \times P(B), \quad \text{for } B \in \mathcal{F}_0^t \times \mathcal{F}_{t+\delta}^\infty.$$

For any $\{B \in \mathcal{F}_0^t \times \mathcal{F}_{t+\delta}^\infty\}$, the set $\Big(\omega : (\omega,\omega) \in B\Big)$ is in $\mathcal{F}$, and since it is also a monotone class, v is well defined. The stochastic process $\phi(t,\omega)$, is said to have Type II φ mixing if

$$\varphi(\delta) = \sup_{t\ge 0} \sup_{A\in\mathcal{F}_0^t\times\mathcal{F}_{t+\delta}^\infty} |\, v_{t,\delta}(A)\,| \xrightarrow[\delta\to\infty]{} 0.$$

Remark on φ mixing: The results we describe hold for Type I mixing as well, both of which were introduced by Volkonskii and Rozanov (1959), since for both types of mixing we have $|v|_{t,\delta}(\Omega \times \Omega) \le 2\varphi(\delta)$.

Let ϵ be a positive number, and consider the system:

$$\begin{aligned} \dot{x}_\epsilon(t,\omega) &= H(x_\epsilon(t,\omega),\omega,t/\epsilon), \\ \dot{y}_\epsilon(t) &= G_\epsilon(y_\epsilon(t),t), \\ x_\epsilon(0,\omega) &= y_\epsilon(0) = x_0 \in R^n. \end{aligned} \tag{2.1}$$

Assume:

1. H is jointly measurable with respect to its three arguments.
2. $G_\epsilon(x,t) = E[H(x(s,\omega),t/\epsilon)]$, and for all i and j

$$\frac{\partial}{\partial x_j} G_i(x,t) \text{ exists, and is continuous in } (x,t).$$

3. For some $T > 0$:

a. There exists a unique solution, $x(t,\omega)$, on $[0,T]$ for almost all ω; and

b. A solution to

$$\frac{\partial}{\partial t} g(t,s,x) = G(g(t,s,x),t), \quad g(s,s,x) = x,$$

exists on $[0,T] \times [0,T] \times R^n$.

The following notations will be used:

1. $H_\epsilon(x_\epsilon(t,\omega),\omega,t) \stackrel{\text{def}}{=} H(x_\epsilon(t,\omega),\omega,t/\epsilon)$
2. $g_s(t,s,x) = (\partial/\partial s)g(t,s,x)$.
3. $g_x(t,s,x) =$ the $n \times n$ matrix with (i,j) component $(\partial/\partial x_j)g_i(t,s,x)$.
4. For $H(x,\omega,\tau)$ define the families of σ-fields $\mathcal{F}_0^t$ and $\mathcal{F}_t^\infty$ such that, for each $t \geq 0$, $\mathcal{F}_0^t$ contains the σ-field generated by

$$\{H(x,\omega,\tau) : 0 \leq \tau \leq t, x \in R^n\},$$

and $\mathcal{F}_t^\infty$ contains the σ-field generated by

$$\{H(x,\omega,\tau) : t \leq \tau < \infty, x \in R^n\}.$$

The relation between the random differential equation and its averaged version for system (2.1) under conditions (1), (2), and (3) is given by:

Lemma (Geman 1977) For any C^1 function $K : R^n \to R^1$ and $t \in [0, T)$:

$$E[K(x(t))] = K(y(t)) + \int_0^t \int_{\Omega\times\Omega} \Big(\frac{\partial}{dx} K(g(t,s,x(s,\omega)))\Big) \cdot H(x(s,\omega),\eta,s) dv_{s,0} ds,$$

provided that

$$\Big(\frac{\partial}{\partial x} K(g(t,s,x(s,\omega)))\Big) \cdot H(x(s,\omega),\eta,s), \quad \text{and}$$
$$\Big(\frac{\partial}{\partial x} K(g(t,s,x(s,\omega)))\Big) \cdot H(x(s,\omega),\omega,s)$$

are absolutely integrable on $\Omega \times \Omega \times [0,T]$, with respect to $dP(\omega)dP(\eta)ds$.

The proof of the lemma is based on the relationship between the initial conditions in time and in space for an ODE, namely: If $g(t,s,x)$ is the function satisfying

$$\frac{\partial}{\partial t} g(t,s,x) = G(g(t,s,x),t)$$

then

$$g_s(t,s,x) = -g_x(t,s,x)G(x,s)$$

for all $t \in [0,\infty)$, $s \in [0,\infty)$, and $x \in R^n$. This follows from the observation that $g(t,s,x)$ is constant along trajectories of the form $(\omega, x(s))$ (cf. Hartman, 1964 chap 5).

Theorem (Geman, 1977) Finite time averaging. Assume also that:

4. There exist continuous functions $B_1(r,t)$, $B_2(r,t)$, and $B_3(r,t)$, such that for all $i, j, k, \tau \geq 0$, and ω:

   a. $| H_i(x,\omega,t,\tau) | \leq B_1(| x |, t)$;

   b. $| (\partial/\partial x_j) H_i(x,\omega,t,\tau) | \leq B_2(| x |, t)$;

   c. $| (\partial^2/\partial x_j \partial x_k) H_i(x,\omega,t,\tau) | \leq B_3(| x |, t)$.

5. $\sup_{\epsilon>0, t\in[0,T]} | y_\epsilon(t) | \leq B_4$ for some B_4 and T.

Then

$$\sup_{t\in[0,T]} \mid x_\epsilon(t) - y_\epsilon(t) \mid \underset{\epsilon\to 0}{\longrightarrow} 0$$

in probability.

3. Averaging on $[0, \infty)$

When averaging on an infinite interval we require that ϵ be a function of t and $\epsilon \searrow 0$, meaning that the mixing rate becomes stronger in time. More specifically, let ϵ be a function of the form $\epsilon(t) = \epsilon_0 \tilde{\epsilon}(t)$ where $\tilde{\epsilon}$ is monotonically decreasing to zero in time.

The above lemma still holds when x, H, g and G are replaced by x_ϵ, H_ϵ, g_ϵ and G_ϵ respectively, and also when ϵ becomes a function of t.

In order for the approximation to hold on $[0, \infty)$ we require that B_1, B_2, B_3 are constants in condition 4 (this will be relaxed later) extend condition 5 to hold for $t \in [0, \infty)$, and add the following relation between the rate of the mixing of H and the convergence of ϵ to zero:

6. $\exists\ \gamma > 0,\ c > 0$, such that $\varphi(\delta) \le \delta^{-\gamma}$, and $\tilde{\epsilon}(t) \le t^{-(\frac{1}{\gamma}+1+c)}$, for a monotone decreasing $\tilde{\epsilon}$.

Theorem 3.1 Assume H_ϵ is of Type II φ mixing, and satisfies condition 1-6, then

$$\lim_{\epsilon_0\to 0} \sup_{t\ge 0} E \mid x_\epsilon(t) - y_\epsilon(t) \mid^2 = 0.$$

Proof: Assume first that t is an integer. Fix ϵ_0 and apply the lemma to the system using $K(x) = \mid x - y_\epsilon(t) \mid^2$:

$$\begin{aligned} E \mid x_\epsilon(t) - y_\epsilon(t) \mid^2 &= \\ &= \mid \int_0^t \int_{\Omega\times\Omega} \Big(\frac{\partial}{\partial x} K(g_\epsilon(t, s, x_\epsilon(s, \omega)))\Big) \cdot H_\epsilon(x_\epsilon(s, \omega), \eta, s) dv_{s,0} ds \mid \\ &\le \sum_{k=1}^{\infty} \mid \int_{k-1}^{k} \int_{\Omega\times\Omega} \Big(\frac{\partial}{\partial x} K(g_\epsilon(t, s, x_\epsilon(s, \omega)))\Big) \cdot H_\epsilon(x_\epsilon(s, \omega), \eta, s) dv_{s,0} ds \mid \end{aligned}$$

For any fixed $\delta_k > 0$ (to be chosen later), since each integral is bounded we can write $\forall k$:

$$\int_{k-1}^{k} \int_{\Omega\times\Omega} \Big(\frac{\partial}{\partial x} K(g_\epsilon(t,s,x_\epsilon(s,\omega)))\Big) \cdot H_\epsilon(x_\epsilon(s,\omega),\eta,s) dv_{s,0} ds =$$

$$\begin{aligned}
I \qquad & = \int_{k-1}^{k-1+\delta_k} \int_{\Omega\times\Omega} \Big(\frac{\partial}{\partial x} K(g_\epsilon(t,s,x_\epsilon(s,\omega)))\Big) \cdot H_\epsilon(x_\epsilon(s,\omega),\eta,s) dv_{s,0} ds \\
II \qquad & + \int_{k-1+\delta_k}^{k} \int_{\Omega\times\Omega} \Big(\frac{\partial}{\partial x} K(g_\epsilon(t,s,x_\epsilon(s-\delta_k,\omega)))\Big) \cdot \\
& \qquad H_\epsilon(x_\epsilon(s-\delta_k,\omega),\eta,s) dv_{s,0} ds \\
& + \int_{k-1+\delta_k}^{k} \int_{\Omega\times\Omega} \{ \Big(\frac{\partial}{\partial x} K(g_\epsilon(t,s,x_\epsilon(s,\omega)))\Big) \cdot H_\epsilon(x_\epsilon(s,\omega),\eta,s) \\
III \qquad & - \Big(\frac{\partial}{\partial x} K(g_\epsilon(t,s,x_\epsilon(s-\delta_k,\omega)))\Big) \cdot H_\epsilon(x_\epsilon(s-\delta_k,\omega),\eta,s) \} dv_{s,0} ds.
\end{aligned}$$

The bounds on x_ϵ and its derivatives, and the smoothness of K imply that I is $O(\delta_k)$. In the second term we can replace $v_{s,0}$ by $v_{s-\delta,\delta}$ since these measures agree on $(\Omega\times\Omega, \mathcal{F}_0^{s-\delta} \times \mathcal{F}_s^{\infty})$, $s \geq \delta$, and since $x_\epsilon(s-\delta,\omega)$ is $\mathcal{F}_0^{s-\delta}$ measurable. Since $v_{t,\delta}$ is the difference of two probability measures, the total variation measure satisfies:

$$|v|_{t,\delta}(\Omega\times\Omega) \leq 2, \text{ and } |v|_{t,\delta}(\Omega\times\Omega) = 2 \sum_{A \in \mathcal{F}_0^t \times \mathcal{F}_{t+\delta}^{\infty}} |v|_{t,\delta}(A),$$

therefore, with Type II (or I) mixing: $|v|_{t,\delta}(\Omega\times\Omega) \leq 2\varphi(\delta)$. Applying this to the second integral and using the above bounds again we get that II is $O\Big(\varphi(\delta_k/\epsilon(k-1))\Big)$. The last term is also $O(\delta_k)$ from the smoothness of H_ϵ and of x_ϵ.

Now choose $\delta_k = \sqrt{\epsilon_0}(k-1)^{-(1+\frac{1}{2}c)}$, $k > 1$, then since $\epsilon(k-1) \leq \epsilon_0(k-1)^{-(\frac{1}{\gamma}+1+c)}$, we get $\delta_k/\epsilon(k-1) \geq \frac{1}{\sqrt{\epsilon_0}}(k-1)^{\frac{1}{\gamma}+\frac{1}{2}c}$. From the condition on φ we have $\varphi(\delta_k/\epsilon(k-1)) \leq \epsilon_0^{\frac{1}{2}\gamma}(k-1)^{-(1+\frac{1}{2}\gamma c)}$. Since $\gamma > 0$, the sum

$$\sum_{k>1} O(\delta_k) + O\Big(\varphi(\delta_k/\epsilon(k-1))\Big) = O\Big(\epsilon_0^{\frac{1}{2}(1+\gamma)}\Big).$$

For the segment of t between two integers, an analogous argument is applied yielding an extra term of the form $O(\epsilon_0^{\frac{\gamma}{2}} + \epsilon_0^{\frac{1}{2}})$. therefore $E \mid x_\epsilon(t) - y_\epsilon(t) \mid^2 = O\Big(\epsilon_0^{\frac{1}{2}(1+\gamma)}\Big)$ uniformly in t.

This implies that

$$\sup_{t\geq 0} E \mid x_\epsilon(t) - y_\epsilon(t) \mid^2 = O\left(\epsilon_0^{\frac{1}{2}(1+\gamma)}\right),$$

$$\lim_{\epsilon_0 \to 0} \sup_{t\geq 0} E \mid x_\epsilon(t) - y_\epsilon(t) \mid^2 = 0.$$

$\diamond$

The following problem is closely related: For fixed ω, let $H(x,\omega,t)$ map $R^n \times R^m \times R^1$ into R^n. Assume that for each x, $H(x,\omega,t)$ is a mixing process, and for each x and t define $G(x,t) = E[H(x,\omega,t)]$. Consider the random equation

$$\dot{x}_{\tilde{\epsilon}}(t,\omega) = \tilde{\epsilon} H(x_{\tilde{\epsilon}}(t,\omega),\omega,t), \quad x_{\tilde{\epsilon}}(0,\omega) = x_0, \tag{3.1}$$

with its averaged equation

$$\dot{y}_{\tilde{\epsilon}}(t) = \tilde{\epsilon} G(y_{\tilde{\epsilon}}(t),t), \quad y_{\tilde{\epsilon}}(0) = x_0. \tag{3.2}$$

For equation (3.2) condition 6 becomes:

6$'$. $\exists \gamma > 0$, such that

i) $\varphi(\delta) < \delta^{-\gamma}$,

ii) $\tilde{\epsilon}(t) = \epsilon_0 r(t) t^{-\rho}$, for $\rho = \frac{1+c}{2+c}$, $c > \frac{1}{\gamma}$, and $\forall t$: $0 < c_1 \leq r(t) < c_2$.

Theorem 3.2 Under the assumptions of theorem (3.1) and (6$'$);

$$\lim_{\epsilon_0 \to 0} \sup_{t\geq 0} E \mid x_{\tilde{\epsilon}(t)} - y_{\tilde{\epsilon}(t)} \mid^2 = 0.$$

Proof: Apply the change of variables: $t = \frac{1}{\epsilon_0}\tau^{2+c}$, $dt = \frac{1}{\epsilon_0}(2+c)\tau^{1+c}d\tau$, to equation (3.1):

$$\begin{aligned}\dot{x}_\epsilon(\tau,\omega) &= \tau^{-\rho(2+c)} r(\tau^{2+c}) H_\epsilon\left(x_\epsilon, \omega, \tau^{2+c}/\epsilon_0\right)(2+c)\tau^{1+c} \\ &= r(\tau^{2+c}) H_\epsilon(x_\epsilon, \omega, \tau/\epsilon(\tau)),\end{aligned}$$

for $\epsilon(\tau) = \epsilon_0 \tau^{-(1+c)}$. Now observe that ϵ satisfies condition (6) in theorem (3.1), which gives the desired result. $\diamond$

As can be seen from the proof, ρ has to satisfy the conditions $\frac{1}{2} < \rho \leq 1$, and $\tilde{\epsilon}(t)$ has to be greater than t^{-1} so that $r(t) \geq c_0 > 0$, which allows the invocation of the previous theorem. It follows that if $\tilde{\epsilon}(t) = t^{-1}$, a convergence is assured for any Type II mixing. Obviously, ρ may be larger than 1 since $\tilde{\epsilon}$ may be split into two functions, one bounded and the other satisfying the conditions of the theorem. The same argument holds for $r(t)$, however, it is clear that one would like $\tilde{\epsilon}$ to go as slow as possible to zero, since then if the averaged version has a limit, the convergence rate of both equations to that limit is inversely proportional to ρ.

It is possible to extend the theory to the cases where the partial derivatives of H have a polynomial growth it time. Then ϵ has to decrease faster so that the above integrals may still be controlled. We get the following theorem:

Theorem 3.3 Assume that B_1, B_2, B_3, and B_4 are bounded by t^α for some $\alpha \geq 0$ in condition 4 of theorem 3.1, and replace condition 6 with the following:

6. $\exists\, \gamma > 0,\ c > \frac{1}{\gamma}$, such that $\varphi(\delta) \leq \delta^{-\gamma}$, and $\tilde{\epsilon}(t) \leq t^{-(1+c+3\alpha)}$, for a monotone decreasing $\tilde{\epsilon}$. Then

$$\lim_{\epsilon_0 \to 0} \sup_{t \geq 0} E \mid x_\epsilon(t) - y_\epsilon(t) \mid^2 = 0.$$

Proof: When applying the lemma as before we get the following:

$$I = \sum_k O(\delta_k)(k-1)^{2\alpha}$$

$$II = \sum_k O(\varphi(\delta_k/\epsilon(k-1))k^{2\alpha}$$

$$III = \sum_k O(\delta_k)k^{3\alpha}.$$

Now chose $\delta_k = \sqrt{\epsilon}_0(k-1)^{-(1+\frac{1}{2}(c-\frac{1}{\gamma})+3\alpha)}$, then since $\epsilon(t) \leq t^{-(1+c+3\alpha)}$, we get just as before $\delta_k/\epsilon(k-1) \geq \frac{1}{\sqrt{\epsilon_0}}(k-1)^{\frac{1}{2}(c-\frac{1}{\gamma})}$. The rest of the proof follows exactly as before. $\diamond$

Extending theorem 3.2 to the case where the partial spatial derivatives are bounded by a polynomial in t is done by absorbing the growth of H into ϵ, which gives the following corollary:

Corollary 3.4 Assume that B_1, B_2, B_3 and B_4 are bounded by t^α for some $\alpha \geq 0$ in condition 4 of theorem 3.1, and replace condition 6 in theorem 3.2 with the following:

6′. $\exists \gamma > 0$, such that

i) $\varphi(\delta) < \delta^{-\gamma}$,

ii) $\tilde{\epsilon}(t) = \epsilon_0 r(t) t^{-(\alpha+\rho)}$, for $\rho = \frac{1+c}{2+c}$, $c > \frac{1}{\gamma}$, and $\forall t$: $0 < c_1 \leq r(t) < c_2$. Then

$$\lim_{\epsilon_0 \to 0} \sup_{t \geq 0} E \mid x_{\tilde{\epsilon}(t)} - y_{\tilde{\epsilon}(t)} \mid^2 = 0.$$

An important observation has to be made here: If the deterministic version represents a converging trajectory, e.g., if the equation represents a gradient descent, then as long as $\tilde{\epsilon}(t) \geq t^{-1}$, the deterministic version will still converge to a true local minimum, however if $\tilde{\epsilon}(t) < t^{-1}$, then $\int_0^\infty \tilde{\epsilon}(\tau) < \infty$, and so the convergence of the deterministic equation is not assured, which implies that the convergence of the stochastic version to a true local minimum is not granted.

4. An application to the synaptic modification equations of a BCM neuron

In this section, we apply the theorem to a random differential equation representing the low governing synaptic weight modification in the BCM theory for learning and memory in neurons, Bienenstock et al. (1982). We start with a short review on the notations and definitions of BCM theory, a more thorough review can be found in Intrator (1990), and the references therein.

Consider a neuron whose input is the vector $x = (x_1, \ldots, x_N)$, has a synaptic-weight vector $m = (m_1, \ldots, m_N)$, both in R^N, and activity (in the linear region) $c = x \cdot m$. The input x is assumed to be a stochastic process of Type II φ mixing, bounded, and piecewise constant. Let $\Theta_m = E[(x \cdot m)^2]$, $\phi(c, \Theta_m) = c^2 - \frac{4}{3}c\Theta_m$. c represents the linear projection of x onto m, and we seek an optimal projection in some sense.

The BCM synaptic modification equations are given by:

$$\dot{m} = \mu(t)\phi(x \cdot m, \Theta_m)x, \quad m(0) = m_0, \tag{4.1}$$

their averaged version is given by:

$$\dot{\bar{m}} = \mu(t)E\Big[\phi(x \cdot \bar{m}, \Theta_{\bar{m}})x\Big], \quad \bar{m}(0) = m_0. \tag{4.2}$$

$\mu(t)$ is a global modulator which is assumed to take into account all the global factors affecting the cell, e.g., the beginning or end of the critical period, or state of arousal (Bear and Cooper, 1988).

Equation (4.2) is shown to be a dimensionality reduction method based on a cost function that favors directions m for which the distribution of the inputs is different from normal by means of skewness (Intrator, 1990).

Our aim is to show the convergence of the stochastic differential equation. This will be done in two step; First we show that the averaged deterministic equation converges, and then we use theorem 3.2 to show the convergence of the random differential equation to its averaged deterministic equation.

The convergence of the deterministic equation

Without loss of generality, we may assume that the random process x is in the unit ball in R^N, and $\mathrm{Var}(x \cdot m) \geq \lambda \| m \|^2 > 0$, which simply says that x does not lie in a subspace or a manifold of R^N. Since we are interested in dimensionality reduction, we can always reduce a-priori the dimensionality of x so that it will span R^N for some N. When the theory is applied to a finite value random vector, $x_1, \ldots, x_n$, we can restrict m to be in the span of $x_1, \ldots, x_n$.

When we multiply both sides of the above equation by $\bar{m}_\mu$, assuming none of its components is zero, we get:

$$\begin{aligned}\frac{1}{2} \parallel \dot{\bar{m}}_\mu \parallel &= E[(x \cdot \bar{m}_\mu)^3] - \frac{4}{3}E^2[(x \cdot \bar{m}_\mu)^2] \\ &\leq \parallel \bar{m}_\mu \parallel^3 - \frac{4}{3}\mathrm{Var}^2(x \cdot \bar{m}_\mu) \\ &\leq \parallel \bar{m}_\mu \parallel^3 - \frac{4}{3}\lambda^2 \parallel \bar{m}_\mu \parallel^4 \\ &= \parallel \bar{m}_\mu \parallel^3 \{1 - \frac{4}{3}\lambda^2 \parallel \bar{m}_\mu \parallel\},\end{aligned}$$

which implies that $\parallel \bar{m}_\mu \parallel \leq \frac{3}{4\lambda^2}$. ◇

Using this fact we can now show the convergence of $\bar{m}_\mu$. We observe that $\dot{\bar{m}}_\mu = -\nabla R$, where $R(\bar{m}_\mu) = -\frac{\mu}{3}\{E[(x \cdot \bar{m}_\mu)^3] - E^2[(x \cdot \bar{m}_\mu)^2]\}$ is the risk. R is bounded from below since $\parallel \bar{m}_\mu \parallel$ is bounded, therefore $\bar{m}_\mu$ converges to a local minimum of R. ◇

The convergence of the stochastic equation

Claim Under the above conditions $m_\mu(t)$ converges in L^2 to a local minimum of the risk.

Proof: The calculation above implies that $\bar{m}_\mu$ is bounded for (almost) every μ.

In our case B_1, B_2, B_3 and B_4 are independent of t or m_μ, therefore, if we replace $\epsilon(t)$ by $\mu(t)$ and apply theorem 3.2, we get

$$\sup_{t \geq 0} E|m_\mu(t) - \bar{m}_\mu(t)|^2 \underset{\mu_0 \to 0}{\longrightarrow} 0.$$

$\bar{m}_\mu$ the solution to the deterministic equation will converge to the same local minimum $\tilde{y}$, $\forall \mu$ if $\mu_0 < C$, for some positive constant C. therefore we can choose $\tilde{T}$ for which $|\bar{m}_\mu(t) - \tilde{y}| < \frac{\delta}{2}$, $\mu_0 < C$, $t \geq \tilde{T}$, then for $t \geq \tilde{T}$ we have:

$$|m_\mu(t) - \tilde{y}| \leq |m_\mu(t) - \bar{m}_\mu(t)| + |\bar{m}_\mu(t) - \tilde{y}| \leq |m_\mu(t) - \bar{m}_\mu(t)| + \frac{\delta}{2},$$

$$\Rightarrow \quad \sup_{t' \geq \tilde{T}} E|m_\mu(t) - \tilde{y}| \leq \sup_{t' \geq \tilde{T}} E|m_\mu(t) - \bar{m}_\mu(t)| + \frac{\delta}{2} \underset{\mu_0 \to 0}{\longrightarrow} \frac{\delta}{2}.$$

δ is arbitrary, which implies that

$$E|m_\mu(t) - \tilde{y}| \underset{\mu \to 0}{\longrightarrow} 0$$

$\diamond$

5. Summary

It has been shown that under mild conditions, the equations $\dot{x}_\epsilon = \epsilon H(x, \omega, t)$, and $\dot{y}_\epsilon = \epsilon G(y, t)$ where $G(x, t) = E[H(x, \omega, t)]$, have close trajectories in the infinite interval when $\epsilon(t) \leq t^{-\frac{1}{2}}$. The result may be computationally useful, and as has been shown in the example, may assist in the analysis of the random differential equation.

Acknowledgments It is a pleasure for me to thank Stuart Geman for suggesting and guiding me through this problem, and to Yali Amit for many helpful remarks. Research was funded in part by the Office of Naval Research, the Army Research Office, the Air Force office of Scientific Research, and the National Science Foundation.

References

Bear M. F., and L. N Cooper (1988) Molecular Mechanisms for Synaptic Modification in the Visual Cortex: Interaction between Theory and Experiment. In *Neuroscience and Connectionist Theory,* M. Gluck and D. Rumelhart, eds.

R. Cogburn, and R. Hersh, Two limit theorems for random differential equations, *Indiana Univ. Math. J.*, 22, (1973) 1067-1089.

P. Dupuis, and H. J. Kushner, Stochastic Approximation and Large Deviations: General Results for w.p.l. Convergence. Tech. Rept. LCDS/CCS No. 87-21, Division of Applied Math., Brown Univ., Providence, RI, 1987.

S. Geman, Averaging for Random Differential Equations. In Approximate Solution of Random Equations, Ed. A. T. Bharucha-Reid North Holland NY, (1977) pp. 49-85.

S. Geman, Some Averaging and Stability Results for Random Differential Equations. *SIAM J. Appl. Math* 36:1, (1979) pp.86-105

P. Hartman, *Ordinary Differential Equations* John Wiley, New York (1964).

N. Intrator, A Neural Network For Feature Extraction. In D. S. Touretzky (ed.), *Advances in Neural Information Processing Systems 2.* San Maeto, CA, (1990) Morgan Kaufmann.

R. Z. Khasminskii, On stochastic process defined by differential equations with a small parameter. *Theory Prob. Application*, 11 (1966), pp. 211-228.

R. Z. Khasminskii, A limit Theorem for Solutions of Differential Equations with Random Right-Hand Side. *Theory Prob. Application*, 11 (1966), pp. 390-406

H. J. Kushner, and D. S. Clark, *Stochastic Approximation methods for Constrained and Unconstrained Systems.* Springer-Verlag.

L. Ljung, Strong Convergence of a Stochastic Approximation Algorithm. *Ann. Statist.*, 6 (1978), pp. 680-696

G. C. Papanicolaou, and W. Kohler Asymptotic theory of mixing stochastic ordinary differential equations. *Comm. Pure Appl. Math. 27,* (1974) 641-668.

R. L. Stratonovich Topics in the Theory of Random Noise. Vols. 1,2, Gordon & Breach, (1963) New York.

V. A. Volkonskii and Yu. A. Rozanov, Some limit theorems for random functions I. *Theory Prob. Appl.*, 4 (1959), pp. 178-197

I. Vrkoc Extension of the averaging method to stochastic equations. *Chekh. Mat. Zh.* 16, (1966) 518-540.

B. S. White Some limit theorems for stochastic delay equations. *Comm. Pure Appl. Math* 29, (1976) 113-141.